DOCUMENT RESUME

ED 171 776 TM 009 234

AUTHOR Smith, Richard M.; Mitchell, Virginia P.
TITLE Improving the Predictive Ability of Placement Tests Using the Rasch Model for Scoring.
PUB DATE Mar 79
NOTE 28p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, California, April, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *College Placement; Complexity Level; Computer Programs; *Goodness of Fit; Grade Prediction; Higher Education; Item Analysis; Mathematical Models; *Predictive Ability (Testing); *Scoring Formulas; Statistical Studies; *Student Ability; Student Placement; Tables (Data); Test Items; *Test Validity
IDENTIFIERS Comparative Guidance and Placement Program; *Rasch Model

ABSTRACT

To improve the accuracy of college placement, Rasch scoring and person-fit statistics on the Comparative Guidance and Placement test (CGP) was compared to the traditional right-only scoring. Correlations were calculated between English and mathematics course grades and scores of 1,448 entering freshmen on the reading, writing, and mathematics sections of the CGP. Results of correlating the three score estimates (right only, Rasch, and person-fit) with final course grades were mixed--it was impossible to identify any systematic differences among the correlations. Before disregarding the Rasch model, it should be recognized that the effect of both traditional and Rasch scoring is mashed by the fact that course grade is not an interval scale; discriminant analysis could be a solution to this problem. In defense of the model, the person-fit analysis was sensitive enough to detect even mild cases of misfit and to recommend score corrections. The BICAL computer program routine for item analysis and recalibration was recommended for test revision when inappropriate items could not be removed from standardized tests. Future studies should develop a test-revising algorithm for the PANAL computer program to provide a completely automated system of misfit identification, test revision, and ability re-estimation. (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED171776

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

IMPROVING THE PREDICTIVE ABILITY OF PLACEMENT TESTS USING THE RASCH MODEL FOR SCORING

Richard M. Smith Virginia P. Mitchell

Mercer County Community College
Trenton, New Jersey

March 1979

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Richard M. Smith

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Paper presented at the annual meeting of the
National Council on Measurement in Education
San Francisco, April, 1979

IM009 234

ABSTRACT

One method of assessing the success of a placement test is to correlate the results of that particular test with some external criterion. It follows, therefore, that for any given placement test, the scoring model which produces the highest correlation between test score and some measure of student success in a course would be the most successful for initial placement. The Rasch Model provides two techniques for modifying the "rights only" scoring technique. The first is the linear ability estimates utilized by the model, and the second is the person-fit statistics which may be derived from the model.

This paper looks at the effect of Rasch scoring and person-fit statistics on the Comparative Guidance and Placement Test--Reading, Writing, and Mathematics sections--and compares these two scoring techniques with "rights only" scoring.

Introduction: The problem of inaccurate student placement has plagued Mercer County Community College for ten years. In addition to the academic difficulties created for the misplaced students, the rest of the students, and the faculty in the classroom, it is thought that the frustration experienced by those students who are placed into courses for which they lack the necessary prerequisite skills and knowledge contributes to the high rate of student attrition at Mercer.

For the past ten years, placement of students into certain English and mathematics courses at the College has been accomplished by the use of scores from the Comparative Guidance and Placement (CGP) test published by the Educational Testing Service. The "rights only" raw scores of the reading and writing portions of the test, when standarized, were found to have correlations with grades in the three primary composition courses ranging from .094 to .361, while the mathematics converted scores correlated more strongly with grades in mathematics courses, .198 to .389, as shown in Table 1. These correlations were based upon the CGP scores and subsequent grades of 1448 entering freshmen from the Fall, 1977, semester but were not dissimilar to those obtained in other studies.

While in some cases nearly 15% of the variance in course grades could be explained by these CGP scores, usually the percentage was nearer to 5%. This is disturbing in itself, as it suggests that this placement system is only marginally helpful in determining which English and mathematics courses students should take. In addition, however, student rates of success in these courses, i.e., the percentage of students earning a grade as opposed to all those attempting the courses, ranged from a low of .31 for elementary algebra to a high of .81 for basic composition. Faculty complaints concerning improper student placement were understandably frequent. While it is known

that many variables affect the determination of student grades and that the inclusion of these variables in the prediction equation would improve the correlation, it seemed logical to continue to explore all available means of selecting students, either for remediation in English and mathematics or for beginning or intermediate college-level work in these subjects.

Possible Solution: It was decided, since the CGP was to remain in use as the primary placement instrument for several years, that one of the first avenues of investigation for improving the placement system should be to employ a different scoring and analysis model for this test. The current scoring model consisted of "rights only" standardized scores. It was the belief of the authors that one of the factors which would affect incorrect placement would be the erronious estimation of the student's ability based on the "rights only" model. This model makes it difficult to identify a variety of measurement problems such as guessing, students' difficulties with different item types, problems with timing, the tendency of good students to misread very easy items, etc., because it cannot predict or estimate quantitatively the probability of success for a student with a given ability on a given item. The person-analysis programs, "PANAL," (Mead, 1978) provide a means of identifying these various types of person misfit. The estimated probability of a given student's success on a given item, when compared with that student's actual success, will allow for a test of goodness-of-fit. This fit statistic may be calculated over any desired grouping of items to check for these measurement problems. The prediction of a student's success on an item, a requirement for "PANAL," is easily obtained via the one parameter logistic model as proposed by Rasch.

The derivation of the Rasch model (Rasch, 1960) is based on the simple propositions that: students are more likely to answer easy items correctly than difficult items; and that students with higher abilities are more likely

to get any given item right than are students with lower abilities. These seemingly obvious statements have a clear connection with the problem of scoring placement tests, especially when the issue is placement into a sequence of courses which require increasingly developed skills within a given area. For example, if students were being tested for placement into a sequence of mathematics courses beginning with the most basic course and leading up to college-level mathematics, it would stand to reason that the lower ability students or those who had trouble answering the easiest items on the placement instrument would be placed into the first course, middle ability students into the next course, and so on. A scoring model which was developed explicitly from these propositions and provides a linear scoring metric in addition to the capacity to offer individual respondent analyses would offer a greater opportunity of correctly identifying which students should be placed into the available courses than would a "rights only" model.

Procedures: The data for this study included the CGP scores of 1448 entering freshman in the Fall, 1977, semester. These tests were administered to all incoming freshman in the period of January, 1977, to August, 1977, prior to admission to the college. Only the initial English and mathematics courses taken by each student were included in these calculations. These courses may have been taken either in the Fall, 1977, or Spring, 1978, semester. For the purpose of this study, only the grades of "A", "B", "C", "D", and "NC" were included. Two sets of correlation coefficients were calculated, one using test scores and the grades "A", "B", "C", "D", "NC", and the other from scores and a dichotomous pass/fail variable formed by collapsing across "A", "B", "C", and "D". The CGP test scores used in this study were converted scores derived by

applying the "rights only" score to the ETS score conversion tables. All course placement was on the basis of these test scores according to the criteria listed in the Appendix.

The initial calibration of items on the six CGP tests and estimates of student abilities were obtained through the use of "BICAL" (Wright and Mead, 1976), a FORTRAN computer program which calculates item difficulties and ability estimates for the Rasch model in a common metric known as "logits" (Wright, 1977).

Wright's contention is that this system of measurement is linear whereas raw score transformations are non-linear. Simply stated, the difference between achieving 50 and 51 questions right on a test of 100 items represents a change in ability which is not the same as the change in ability encountered when going from 99 to 100 items since the score metric in non-linear. However, the difference between an ability estimate of 0.0 and 0.5 logits is exactly the same as the ability difference between 3.0 and 3.5 logits.

This linearization of the raw score is accomplished by compressing differences in ability at the center of the raw score distribution and extending the differences of ability for the upper and lower limits of the score range, as in Table 2. In this example the ability difference between 20 items correct and 21 items correct is .11 logits, while the ability difference between 38 items correct and 39 items correct is .72 logits.

The benefit of this procedure was that, since the same metric was used for estimates of both item difficulty and student ability, the probability of a student with a given ability getting an item with a given difficulty correct could be calculated. The final step of this procedure was the calculation of item "t" statistics to demonstrate how well the probabilities derived from the model fit the actual response patterns of the students.

Through an examination of the item "t" statistics two kinds of inappropriate items were eliminated from the tests. Items exhibiting the quality of "overfit", i.e., those items for which expected probabilities of low ability students exceeded the actual percentage of correct responses for that group while the expected probabilities of high ability students underestimated the actual percentage of correct responses, were deleted. Items exhibiting the quality of "underfit," or those items for which the low ability students did better than expected and the high ability students did worse than expected, were also deleted.

On the basis of this analysis of item misfit all misfitting items were excluded and the tests were reanalyzed. Since the calibration process is "sample-freed" and the ability estimates "item-freed" (Waller, 1976), the removal of misfitting items would not affect the recalculation of the ability estimates. The recalculation produced the so-called "revised Rasch ability estimates."

The individual person-fit statistics were then calculated from the original item calibration and the individual item responses using the PANAL program. All individuals having a total misfit statistic greater than two were examined on three item groupings of misfit. These three types of misfit were: misfit resulting from item difficulty; misfit resulting from the item's position on the test; and misfit resulting from the subject content of the items. Misfit due to item difficulty was also investigated to detect changes in a student's response pattern as items became either too difficult or too easy. A between group "t" value greater than 2 was used to select cases of misfit. In instances where misfit occurred, the misfit was corrected by limiting the range of item difficulties so that no item was more than one logit above the student's ability or two logits below the student's ability, i.e., $-2 \leq (B_\gamma - \xi_i) \leq 1$ where B_γ is the individual ability estimate and ξ_i is the item difficulty estimate.

Waller (1976) relates that latent trait theory suggests ability estimates from any subset of items may be calculated on the same ability scale as the total test. This seems in keeping with the "item freed" claims of the Rasch Model and should allow for accurate estimates of items where the individual's response pattern fits the model.

Misfit due to item placement was used to detect problems relating to start-up and the speededness of the test. Students who experienced difficulty early in the test on easy items, or students who exhibited guessing behavior on the latter parts of the test, were identified. The appropriate adjustments for test length were made. That is, ability estimates were calculated only on the basis of item responses which fit the Rasch Model.

On the basis of the two types of misfit discussed, the appropriate modifications were made in the test, and the estimates of person ability were recalculated for all misfitting students. Correlation coefficients were then calculated for the two measures of course success and the four measures of individual ability. These four ability estimates were:

A. Recommended raw score conversions.

B. Rasch ability estimates based on the total test.

C. Rasch ability estimates based on the test excluding misfitting items.

D. Rasch ability estimates based on the total test but correlated for person misfit.

Results: The original Rasch calibration, using "BICAL", of the six placement tests (Reading, Sentences, Total English, Math Part 1, Math Part 2, Total Math) provided six log ability estimates for each of the students. This calibration also indicated that there were between 5 and 13 misfitting items on each test. These items exhibited both "overfit" and "underfit" as calculated by the fit mean square statistic utilized in the BICAL pro-

gram. The items were chosen for deletion if the between group fit mean square and total fit mean square were more than five standard deviations from the expected value of 1. The item deletion procedure followed here was harsh, in that items were omitted solely on the basis of Rasch statistics, and neither item content nor format was reviewed to determine the cause of misfit. If the results of this study were to be used for actual course placement, the criteria for omitting items would probably have been relaxed in order to retain more items, thus decreasing the standard error of measurement for the test. This would improve the accuracy with which the test score characterizes the individual's true ability.

Following the identification and deletion of the misfitting items, the tests were again calibrated and a second set of Rasch log ability estimates were calculated. These estimates based on the shorter revised tests will be called the "revised Rasch ability estimates". A review of the fit mean square statistics revealed that, despite the deletion of misfitting items, there still was measurement "noise", i.e. items which had acceptable fit mean square statistics in the first calibration had now become misfitting. This probably indicated a lack of unidimensionality in the scale or trait defined by the items and not an inherent problem with the items. The change in the fit mean square statistic supports the belief that the original deletion criteria were too stringent, that many of the omitted items were acceptable, and that, given the original set of items, little can be done in the way of improving item fit without rewriting portions of the tests.

The results of correlating these three sets of score estimates (standard score, Rasch ability estimates, and revised Rasch estimates) with the final course grades for three levels of English and two levels of mathematics were mixed as shown in Tables 2, 3, 4, and 5. It was impossible to identify

any systematic differences among the correlations. For English, the differences in correlations between standard scores and grades vs. Rasch ability estimates and grades for the two tests and the total test ranged from - .064 to + .016. For math these same differences ranged for - .032 to + .092. The differences in the correlations between the Rasch ability estimates and the revised Rasch ability estimates exhibited the same variability. These values ranged from + .037 to - .062 for English course grade prediction and from + .032 to - .068 for mathematics. These differences were calculated from correlation values which ranged from - .003 to .409, when grades were calculated as a polychotomous variable ("A", "B", "C", "D", "NC"), and from - .071 to .336, when grades were calculated as a dichotomous pass/fail variable.

The responses of the original calibration sample to the Sentences test was then analyzed using the "PANAL" system of programs. A total of 38 persons were found to have a total "t" fit statistic greater than 2. These values of "t" should occur infrequently by chance in this distribution and these cases should, therefore, be investigated. This represents 1.9 percent of the total population tested. Using the between-group "t" statistic calculated for four sets of items grouped by Rasch item difficulty, 46 persons were found to have a "t" value greater than 2. This represents 2.3% of the total population tested. Eleven of the persons in the former group also fell into the latter, leaving a total of 73 individuals who were identified as misfitting, i.e., exhibiting unusual response behavior.

Revised ability estimates for these individuals were then calculated using several criteria for identifying areas of misfit, including the assumption that item difficulties should fall into a range no greater than one logit above and two logits below the student's ability. Considerations of guessing and "test sloppiness", i.e., the tendency of respondents to miss

items considerably below their estimated abilities, were also utilized to attempt to derive more accurate ability estimates for those individuals exhibiting misfit.

The correlation of these revised person-scores with final English course grade was of marginal value. Of the 73 individuals identified as misfitting, only 30 had enrolled in English courses. Of these 30, 11 enrolled in ES100, 14 enrolled in EG101, and five enrolled in EG102. The results of these correlations are listed in Table 8. These results show no consistant differences between the predictive ability of the original Rasch ability estimates and the revised person-ability estimates. The correlations for these subgroups were also widely different from those for the original groups from which they were drawn.

Conclusions: There are several conclusions which can be drawn from the results of this study. Most of these conclusions are, in fact, double edged. From a conservative point of view, the lack of significant improvement in the predictive ability of the test using the Rasch ability estimates or the revised Rasch ability estimates could be considered as a confirmation of the validity ofthe traditional approach to test analysis and score reporting and, therefore, a reason for not employing the Rasch model.

There are several factors which should be considered before drawing such conclusions from this study, however. The Rasch ability estimate is a monotonic function of the raw score. The tendency of the Rasch Model to linearize the scoring metric will not affect the rank ordering of the students' scores. Without any change in the rank ordering there is little chance of significantly altering the correlation of scores with any external criterion. Therefore, the similarity of the correlation values obtained using the standard score reporting system and the Rasch ability estimates to predict final grade is not totally unexpected.

The true nature of the effect of both score reporting systems is masked by the fact that the course grade is not an interval scale. This tends to confound the calculation of the correlation, while the restricted range of test scores resulting from the use of these scores for course placement would tend to depress the correlation. In addition, other studies indicate that, for these courses, grades are much more highly correlated with biographical data than with test results. When biographical factors are considered, "R" values for predicting course grade can be raised substantially, as shown in Table 9.

The practical utility of the Rasch Model lies in its ability to describe a student's ability in terms of the types of behaviors which he or she is capable of performing, and, consequently, in its ability to calculate the probability of a student correctly answering any given item.

The inability of the revised Rasch ability estimates systematically to produce better correlations than the original Rasch estimates leads to somewhat different conclusions. The item deletion strategy employed was quite severe, leading to the elimination of between 13% and 37% of the test items for each test. With reductions of this magnitude in test length and the resulting increase in the standard error, the lack of any systematic differences in grade predictive ability would seem to imply that the shorter tests were no less effective in predicting student success in courses than the original versions.

The utilization of the person-analysis and the heuristic application of score revision guidelines based upon this analysis provided an excellent means of verifying that the behaviors exhibited by individuals in the testing situation were in keeping with the expectations of the Rasch Model. The identification of only 73 individuals out of 2024 who exhibited misfit on the two criteria utilized was encouraging in that the vast majority of students took the

testing seriously, thereby following the model, and that the person-analysis was sensitive enough to detect even mild cases of misfit. Instances in which students failed to respond to as few as five items were detected as misfitting.

In 72 of the 73 cases identified, it was possible to identify the source of the misfit and to recommend score corrections. It should be understood, however, that only two forms of misfit were investigated. These were total misfit based upon the student's interaction with all of the test items, and the between-group misfit based on the student's interaction with four ordered item difficulty subgroups. There were many other possible sources of misfit which could have been investigated, such as item type, item sequence, content subscales, and the cognitive level of items. Several of these other sources of misfit were calculated for this sample, although they were not reported in the results of this limited study. If all of the sources of misfit were included the total number of misfit would have approached 10% of the total sample.

The description of the causes of unusual response patterns was greatly enhanced by the examination of as many sources of misfit as possible, using the "PANAL" Programs. It was often the case that problems in the identification of the underlying cause of a student's unexpected behavior based solely on item difficulty encouraged a reanalysis of the data using other factors such as those mentioned above. Each attempt at reanalysis resulted in the identification of previously unselected students who were misfitting on the new factor.

Recommendations: There are two major concerns which the researcher should consider when comparing the predictive abilities of the Rasch Model and the "rights only" scoring model, i.e., the choice of an external criterion and the choice of a statistical technique to demonstrate the difference between the models, should one exist. The problem with the choice of the final course

grade as the external criterion is that it can't be represented by an interval scale of measurement. As mentioned above, this causes a problem when using the Pearson product moment correlation. Also, the final course grade is influenced by many factors other than entering competencies as measured by a placement test. Such factors often have a more important role in influencing final grade, thus strongly affecting the correlations. The solution might either be to select an external criterion more in keeping with the statistical properties of product moment correlations or to choose a more appropriate statistical technique.

The use of discriminant analysis could be a solution to the problem of the non-interval nature of the final course grade variable, particularly if the study was to determine which model was best for classifying students into grade sub-groups. The consideration of this technique should be part of any further attempt to study this subject.

Some points should be made about the implementation of the Rasch Model using the "BICAL" program. The procedures for item calibration and ability estimation allow recalibration with the change of a single control card. In recalibration the item difficulties were found to be extremely stable, thus yielding ability estimates which were on the same scale with the same origin as the original ability estimates.

The program also provides a form of item analysis which is useful in identifying misfitting items. This routine, combined with the recalibration capacity mentioned above, suggests an interesting solution to a difficult problem which is often encountered in mandated testing programs. Since the revised ability estimates are on the same scale as the original estimates, it would be possible to administer a mandated test, calibrate and remove misfitting items, recalibrate the remaining items, and estimate

students' abilities based on the locally defined subtest. This would be particularly useful in any situation in which a test which was designed to meet the needs of a diverse population is being used to test a fairly narrow subpopulation. Since the revised ability estimates are not dependent on any given subset of items, it is possible to include or exclude items from ability estimation without revising the cut-off scores which were determined on the basis of the entire test. It is, therefore, recommended that this technique of test revision be employed in instances where inappropriate items cannot be removed from the standardized testing instruments.

The person-analysis aspect of this study suggests a solution to the day-to-day problem of making decisions about test scores at or near the chance level. This situation is particularly true for mandated testing instruments with item difficulties which cannot be changed. The analysis of low range scores, differentiating between random behavior and minimal knowledge of the material, is a welcome prospect.

In this regard, it is unfortunate that the number of misfitting students on the CGP Sentences test who took English courses was too low to allow any conclusions to be drawn. However, from a heuristic viewpoint, the person-analysis procedure seemed reasonable, workable, and worthwhile. The capacity to determine expected behavior of a student on an item-by-item basis and to calculate departures from expected values over any user-defined subset of items has great practical applicability.

The "PANAL" programs have been shown to be sensitive to relatively minor departures from student behavior as predicted by the model. With a judicious choice of "t" values and subgroups, it is possible to screen large numbers of tests with relative efficiency and select out misfitting cases. The model's predictions of misfit and the authors' internalized conception of misfit usually corresponded quite well. There was little, if any, disagreement with

the cases selected as misfitting by "PANAL" via the "t" statistic, and a review of all other cases indicated that slight adjustments in the "t" value would capture most of the few remaining cases which were misfitting but hadn't been selected by "PANAL".

Given the efficient selection of misfitting cases, the information provided by the "PANAL" program made it possible to define subtests of the total test on which students' abilities could be more accurately estimated with little practice. The authors were able to identify the causes of misfit and define such acceptable subtests with relative ease for 72 of the 73 misfit cases. Only one of the 73 posed any problem, and it is hypothesized that the addition of an item content by item difficulty analysis could have explained these causes.

The utilization of the "item-freed" nature of person ability estimation in the Rasch Model should provide more reliable and defensible ability estimates than do the heuristic total raw score adjustments used in this study. With experience, it should be possible to develop a test-revising algorithm to be included in the "PANAL" program to provide a completely automated system of misfit identification, test revision, and ability re-estimation. It is towards this end that future studies should be directed.

CORRELATION BETWEEN COURSE GRADE AND PLACEMENT TEST RESULTS

Course	Test		
	Reading	Writing *	Math
ES100: Basic Composition	.314	.361	
EG101: English Composition I	.180	.250	
EG102: English Composition II	.094	.221	
MS100: Basic Mathematics			.322
MA109: Elementary Algebra			.198**
MA122: Intermediate Algebra			.316
MA115: Algebra and Trigonometry I			.389
MA103: College Algebra I			.232

* A total score computed by adding CGP Sentences and CGP Reading scores.

** Only "A", "B", and "NC" can be earned in this course.

Table 2

CORRELATION VALUES FOR RIGHTS ONLY SCORING,

RASCH ABILITY ESTIMATES, AND REVISED RASCH ABILITY ESTIMATES WITH COURSE GRADES

	CGP Reading	CGP Sentences	CGP Total	RASCH Reading	RASCH Sentences	RASCH Total	RR Reading	RR Sentences	RR Total
ES100 N=144	.314	.361	.409	.327	.362	.408	.337	.350	.409
EG101 N=392	.180	.250	.259	.190	.259	.263	.203	.254	.278
EG102 N=216	.094	.221	.182	.065	.156	.154	-.003	.188	.129

Table 3

CORRELATION VALUES FOR RIGHTS ONLY SCORING,

RASCH ABILITY ESTIMATES, AND REVISED RASCH ABILITY ESTIMATES WITH COURSE GRADES

	CGP Computation	CGP Applied Arithmetic	CGP Total	RASCH Comp	RASCH A.A	RASCH Total	RR Comp	RR A.A	RR Total
MS100 N=107	.302	.240	.322	.292	.278	.348	.243	.233	.369
MA109 N=202	.241	.120	.198	.198	.212	.255	.168	.244	.244

Table 4

CORRELATION VALUES FOR RIGHTS ONLY SCORING,

RASCH ABILITY ESTIMATES, AND REVISED RASCH ABILITY ESTIMATES WITH A PASS/FAIL MEASURE OFCOURSE SUCCESS

	CGP Reading	CGP Sentences	CGP Total	RASCH Reading	RASCH Sentences	RASCH Total	RR Reading	RR Sentences	RR Total
ES100 N=144	.273	.274	.336	.273	.282	.336	.288	.273	.342
EG101 N=392	.078	.128	.129	.080	.129	.119	.103	.117	.134
EG102 N=216	-.054	.092	.024	-.034	.077	.040	-.071	.103	.037

Table 5

CORRELATION VALUES FOR RIGHTS ONLY SCORING,

RASCH ABILITY ESTIMATES, AND REVISED RASCH ABILITY ESTIMATES WITH PASS/FAIL MEASURE OF COURSE SUCCESS

	CGP Computation	CGP Applied Arithmetic	CGP Total	RASCH Comp	RASCH A.A	RASCH Total	RR Comp	RR A.A	RR Total
MS100 N=107	.200	.143	.209	.203	.163	.218	.183	.103	.157
MA109 N=202	.188	.082	.149	.156	.136	.223	.129	.171	.255

17

Table 6

DIFFERENCES BETWEEN CORRELATIONS FOR VARIOUS SCORING TECHNIQUES

	Reading		Sentences		Total	
	Rights Only vs Rasch Ability	Rasch Ability vs Revised Rasch	Rights Only vs Rasch Ability	Rasch Ability vs Revised Rasch	Rights Only vs Rasch Ability	Rasch Ability vs Revised Rasch
ES100	+.013	+.010	+.001	-.011	.000	+.001
ES100P*	+.001	+.015	+.008	-.009	.000	+.006
EG101	+.009	+.014	+.010	-.004	+.005	+.015
EG101P	+.002	+.023	+.001	-.011	-.005	+.015
EG102	-.029	-.062	-.064	+.032	-.027	-.026
EG102P	-.020	+.037	-.015	+.026	+.016	-.003

*P - Indicates that the variable was collapsed to a dichotomous pass/fail.

Table 7

	Computation		Applied Arithmetic		Total	
	Rights Only vs Rasch Ability	Rasch Ability vs Revised Rasch	Rights Only vs Rasch Ability	Rasch Ability vs Revised Rasch	Rights Only vs Rasch Ability	Rasch Ability vs Revised Rasch
MS100	-.010	-.050	+.038	-.045	-.001	-.068
MS100P*	+.003	-.020	+.020	-.060	+.009	-.060
MA109	-.043	-.030	+.092	+.032	+.091	+.030
MA109P	-.032	-.027	+.054	+.005	+.074	+.032

*P - Indicates that the variable was collapsed to a dichotomous pass/fail.

Table 8

CORRELATION OF FINAL COURSE GRADE WITH

ORIGINAL RASCH ABILITY AND PERSON REVISED ABILITY FOR MISFITTING INDIVIDUALS

Course	Original Rasch Ability	Person Revised Ability
ES100 (N=11)	.634	.581
EG101 (N=14)	.169	.209
EG102 (N=5)	.488	.500

Table 9

CORRELATION BETWEEN PRE-COURSE INDICATORS

AND FINAL GRADE

Course	Test	Test & Biographic Variables
MS100	.322	.744
MA109	.198	.375
MA103	.232	.739
ES100	.361	.645
EG101	.250	.446
EG102	.221	.446

References

Mead, R. J., PANAL Programs, University of Chicago, 1978.

Rasch, G., Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen, Denmark: Danmarks Paedogogiske Institut, 1960.

Ryan, J. P., and Hamm, D. W., An introduction to the Rasch latent trait psychometric model. Paper presented to a pre-session at the annual conference of the American Educational Research Association, Toronto, 1978.

Waller, M., Ability removing random guessing model. Paper presented at the annual conference of the American Educational Research Association, San Francisco, 1976.

Wright, B. D., Solving measurement problems with the Rasch model. Journal of Educational Measurement, Vol. 14, no. 2, pp. 97-116.

Wright, B. D., The Rasch model for test construction and person measurement. A paper presented to the fifth annual Conference and Exhibition on Measurement and Evaluation, Los Angeles, 1978.

Wright, B. D., and Mead, R. J. BICAL: Calibrating items and scales with the Rasch model. Research Memorandum Number 23A, University of Chicago, 1978.

APPENDIX:

CRITERIA FOR PLACEMENT AND COURSE
GRADE DESCRIPTIONS

ENGLISH PLACEMENT

Course	Score Range*
ES100	76 or Below
Writing Sample	77 - 83
EG101	84 - 102
Writing Sample	103 - 110
EG102	111 - 121
Writing Sample	122 or Above

* A combined score from the Reading and Writing Tests

MATHEMATICS PLACEMENT

Math Test	Math Course	Score Range
C *	MS109	Less than 46
	MA109	46 and Above
D **	MA109	Less than 50
	MA122	Greater than 50
E ***	MA122	Less than 59 with D Grade
	MA103	Less than 59 with C Grade
	MA104	Greater than 60

* Math Test C is composed of two sections: Computation and Applied Arithmetic
** Math Test D is composed of two sections: Computation and Applied Algebra
*** Math Test E is composed of two sections: Elementary Algebra and Inter. Algebra

Course Grade

A - Outstanding Achievement

B - Above Average Achievement

C - Average Achievement

D - Sufficient Achievement

NC - No Credit. Student did not satisfactorily complete the course work or withdrew after the allotted time.

W - Withdrew Officially. No evaluation (applicable through 60% of the course).

An asterisk (*) may be attached to B, C, D, NC letter grades and will indicate an earned grade, subject to improvement within one year.